



Calculer sur la langue mais qu'y comprendre ?

Maxime Amblard

► To cite this version:

Maxime Amblard. Calculer sur la langue mais qu'y comprendre ?. Colloque Cathy Dufour, Nov 2018, Nancy, France. hal-01941846

HAL Id: hal-01941846

<https://inria.hal.science/hal-01941846>

Submitted on 2 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Calculer sur la langue mais qu'y comprendre ?

Colloque Cathy Dufour

Maxime Amblard

14 novembre 2018



Calculer sur la langue

Modéliser formellement le sens

Modéliser stochastiquement le sens

Compositionnalité dans les réseaux de neurones

Conclusion

Calculer sur la langue

1. L'IA s'impose comme un enjeu avec des domaines et des méthodes
2. Le traitement de la langue (**natural language**) peut être inclus dans l'IA

1.

1. L'IA s'impose comme un enjeu avec des domaines et des méthodes
2. Le traitement de la langue (**natural language**) peut être inclus dans l'IA ¹

1. On se souviendra que le traitement de la langue est antérieur à l'IA, mais comme aujourd'hui on utilise les mêmes outils et que le terme IA a gagné, on ne polémiquera pas sur l'inclusion du TAL dans l'IA ou de l'IA dans le TAL.

1. L'IA s'impose comme un enjeu avec des domaines et des méthodes
2. Le traitement de la langue (**natural language**) peut être inclus dans l'IA ¹
3. Volonté de **calculer** sur la langue pour produire des représentations et des outils

1. On se souviendra que le traitement de la langue est antérieur à l'IA, mais comme aujourd'hui on utilise les mêmes outils et que le terme IA a gagné, on ne polémiquera pas sur l'inclusion du TAL dans l'IA ou de l'IA dans le TAL.

1. L'IA s'impose comme un enjeu avec des domaines et des méthodes
2. Le traitement de la langue (**natural language**) peut être inclus dans l'IA ¹
3. Volonté de **calculer** sur la langue pour produire des représentations et des outils
4. Volonté d'y **inclure** les propriétés de la langue (*i.e.* **la linguistique**)

1. On se souviendra que le traitement de la langue est antérieur à l'IA, mais comme aujourd'hui on utilise les mêmes outils et que le terme IA a gagné, on ne polémiquera pas sur l'inclusion du TAL dans l'IA ou de l'IA dans le TAL.

Calculer et inclure la linguistique

Calculer et inclure la linguistique

- Natural Language Processing (NLP)

Traitement Automatique des Langues (TAL)

- Computational Linguistics

Linguistique informatique

Calculer et inclure la linguistique

- Natural Language Processing (NLP)

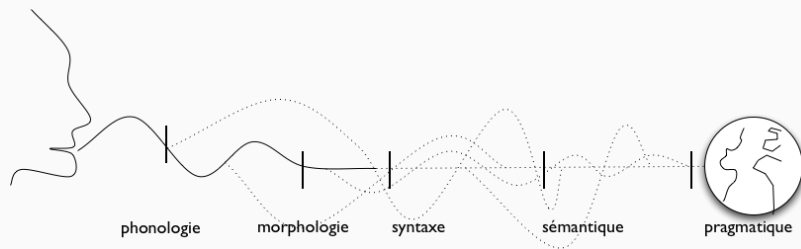
Traitement Automatique des Langues (TAL)

engineering, résolution de tâches, approche expérimentale par l'évaluation

- Computational Linguistics

Linguistique informatique

modèles explicatifs, validation par les données



Modélisation de la langue naturelle :

- Analyse ou génération ?

Modélisation de la langue naturelle :

- Analyse ou génération ?
- Deux paradigmes :
 - Paradigme symbolique (Grammaires, lexiques, systèmes à base de règles)
 - Paradigme stochastique (Machine learning, réseau de neurones)

Où :

- mails, SMS, Twitter, Facebook, commentaires, blogs, articles, pages web, ...

Où :

- mails, SMS, Twitter, Facebook, commentaires, blogs, articles, pages web, ...

Pour quoi faire :

- interface humain-machine, dialogues automatiques
- traduction automatique, analyse d'opinion
- génération de documents, résumé
- Watson !
- recherche d'information, représentation de l'information

Sous-domaines très riches : médical, légal, champs techniques

UNDERSTANDING ONLINE STAR RATINGS:



VS

THE GRINCH MOVIE REVIEW

TIMES OF INDIA

Neil Soans, Nov 7, 2018, 10.30 PM IST

Critic's Rating: ★★☆☆☆ 3.5/5

The Grinch Story: The town of Whoville loves Christmas, and each year their celebrations get bigger and louder. That doesn't go down well with The Grinch, who detests the festive season.

The Grinch Review: If you're familiar with the Dr Seuss children's classic "How the Grinch Stole Christmas!", you'll be familiar with this story as well. How can the

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie
- syntaxe
- sémantique
- pragmatique

- phonologie
- morphologie

Ginette Garcin

écouter

- syntaxe

- sémantique

- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

autossa = dans l'auto
autossaniko = dans mon auto

- syntaxe
- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

la : pronom, nom ou déterminant ?

- syntaxe
- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

Jean regarde (la fille avec un télescope)
Jean regarde (la fille) (avec un télescope)

- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie

- syntaxe

Jean pense que Paul croit que Guillaume dit qu'il conduira la voiture

- sémantique
- pragmatique

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie
- syntaxe
- sémantique
- pragmatique

La belle ferme la porte
étoile : célébrité ou astre ?

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie
- syntaxe
- sémantique
- pragmatique

La pièce est dans le porte-monnaie
Le porte-monnaie est dans la pièce

Modéliser la langue = gérer les ambiguïtés

- phonologie
- morphologie
- syntaxe
- sémantique
- pragmatique

Modéliser formellement le sens

LANGAGE



MONDE

LOGIQUE



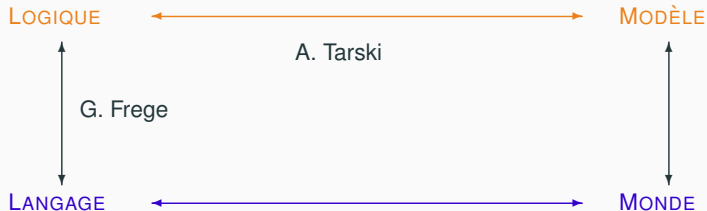
G. Frege

LANGAGE

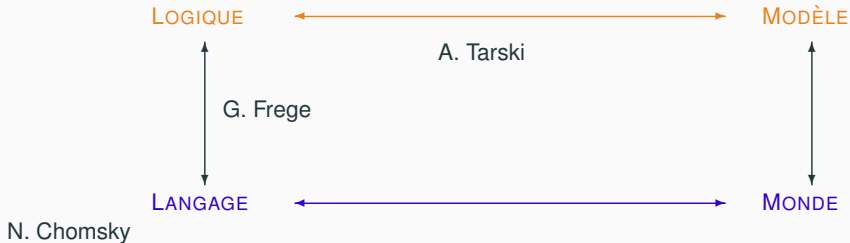


MONDE

Principe de compositionnalité



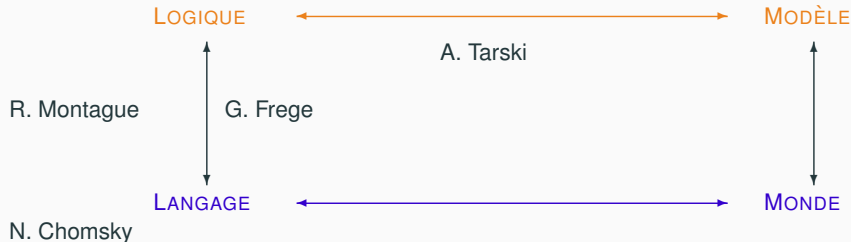
Principe de compositionnalité
Satisfiabilité



Principe de compositionnalité

Satisfiabilité

Linguistique computationnelle



Principe de compositionnalité

Satisfiabilité

Linguistique computationnelle

Sémantique computationnelle

- Semantique

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)
- (1) John loves Mary

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)
- (1) John loves Mary

love(John, Mary)

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)

(1) John loves Mary

love(John, Mary)

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)

(1) John loves Mary

love(John, Mary)

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?

(2) If a farmer owns a donkey, he beats it

- Semantique
- Compositionnalité (Frege)

Le sens d'une expression complexe est la composition du sens de ses parties

Rq : La langue est massivement compositionnelle bien qu'elle ne le soit pas complètement

- Approches logique (Montague)

(1) John loves Mary

love(John, Mary)

- Comment utiliser ces représentations ?
- Utilité de ces représentations ?

(2) If a farmer owns a donkey, he beats it

$\forall x \forall y. ((\text{farmer } x \wedge \text{donkey } y \wedge \text{own } x \ y) \rightarrow \text{beat } x \ y)$

Représentation DRT (KAMP et REYLE 1993) et SDRT (ASHER et LASCARIDES 2003)

Représentation DRT (KAMP et REYLE 1993) et SDRT (ASHER et LASCARIDES 2003)

Old mathematicians never die ; they
just lose some of their functions.

Représentation DRT (KAMP et REYLE 1993) et SDRT (ASHER et LASCA- RIDES 2003)

Old mathematicians never die ; they
just lose some of their functions.

x1	s1	e1	t1	t2	s2
k1 ::					
Theme(s1, x1)					
old(s1)					
mathematician(x1)					
die(e1)					
Patient(e1, x1)					
now(t1)					
e1 \subseteq t2					
t2 = t1					
Manner(e1, s2)					
never(s2)					

x1	x2	x1	x3	e2	t1	t3	s3
k2 ::							
thing(x1)							
thing(x2)							
thing(x1)							
of(x3, x1)							
function(x3)							
of(x2, x3)							
lose(e2)							
Agent(e2, x1)							
Theme(e2, x2)							
now(t1)							
e2 \subseteq t3							
t3 = t1							
Manner(e2, s3)							
just(s3)							

continuation(k1, k2)
parallel(k1, k2)

Représentation DRT (KAMP et REYLE 1993) et SDRT (ASHER et LASCA- RIDES 2003)

Old mathematicians never die ; they
just lose some of their functions.

x1	s1	e1	t1	t2	s2
k1 ::					
Theme(s1, x1)					
old(s1)					
mathematician(x1)					
die(e1)					
Patient(e1, x1)					
now(t1)					
e1 \subseteq t2					
t2 = t1					
Manner(e1, s2)					
never(s2)					

x1	x2	x1	x3	e2	t1	t3	s3
k2 ::							
thing(x1)							
thing(x2)							
thing(x1)							
of(x3, x1)							
function(x3)							
of(x2, x3)							
lose(e2)							
Agent(e2, x1)							
Theme(e2, x2)							
now(t1)							
e2 \subseteq t3							
t3 = t1							
Manner(e2, s3)							
just(s3)							

continuation(k1, k2)
parallel(k1, k2)

An engineer thinks that his equations
are an approximation to reality.
A physicist thinks reality is an
approximation to his equations.
A mathematician doesn't care.

Représentation DRT (KAMP et REYLE 1993) et SDRT (ASHER et LASCA-RIDES 2003)

Old mathematicians never die ; they just lose some of their functions.

x1	s1	e1	t1	t2	s2	x1	x2	x1	x3	e2	t1	t3	s3
k1 ::	Theme(s1, x1)					k2 ::	thing(x1)						
	old(s1)						thing(x2)						
	mathematician(x1)						thing(x1)						
	die(e1)						of(x3, x1)						
	Patient(e1, x1)						function(x3)						
	now(t1)						of(x2, x3)						
	e1 \subseteq t2						lose(e2)						
	t2 = t1						Agent(e2, x1)						
	Manner(e1, s2)						Theme(e2, x2)						
	never(s2)						now(t1)						
							e2 \subseteq t3						
							t3 = t1						
							Manner(e2, s3)						
							just(s3)						
continuation(k1, k2)													
parallel(k1, k2)													

An engineer thinks that his equations are an approximation to reality.
A physicist thinks reality is an approximation to his equations.
A mathematician doesn't care.

x1	p1	e1	t1	t2	x5	e2	p2	t1	t4	x9	
k1 ::	engineer(x1)				k2 ::	physicist(x5)				k3 ::	mathematician(x9)
	p1: x1 x2 x3 x4 s1 t1 t3					think(e2)					- a3 t1 t5
	male(x1)					Theme(e2, x5)					care(e3)
	of(x2, x1)					Theme(x2, p2)					Experience(e3, x9)
	equation(x2)					p2: x6 x7 x5 x8 s2 t1 t5					now(t1)
	approximation(x3)						reality(x6)				a3 \subseteq t5
	reality(x4)						approximation(x7)				t5 = t1
	to(x3, x4)						male(x5)				
	identical(s1)						of(x8, x3)				
	Pivot(s1, x2)						equation(x8)				
	Theme(s1, x3)						to(x7, x8)				
	now(t1)						identical(p2)				
	s1 \subseteq t3						Pivot(s2, x6)				
	t3 = t1						Theme(s2, x7)				
	think(e1)						now(t1)				
	Agent(e1, x1)						s2 \subseteq t5				
	Theme(e1, p1)						t5 = t1				
	now(t1)										
	e1 \subseteq t2						now(t1)				
	t2 = t1						e2 \subseteq t4				
							t4 = t1				
continuation(k1, k2)											
continuation(k2, k3)											

- Nombreuses stratégies pour **modéliser finement** les phénomènes de sens dans les énoncés
- Basées sur la **compositionnalité**

- Nombreuses stratégies pour **modéliser finement** les phénomènes de sens dans les énoncés
- Basées sur la **compositionnalité**

Mais

- Nombreuses stratégies pour **modéliser finement** les phénomènes de sens dans les énoncés
- Basées sur la **compositionnalité**

Mais

- Modéliser (parfaitement) un phénomène particulier **n'est pas comprendre le sens** général de l'énoncé
- Difficulté d'avoir une représentation complète du sens (ce que fait très bien la langue)

- Nombreuses stratégies pour **modéliser finement** les phénomènes de sens dans les énoncés
- Basées sur la **compositionnalité**

Mais

- Modéliser (parfaitement) un phénomène particulier **n'est pas comprendre le sens** général de l'énoncé
- Difficulté d'avoir une représentation complète du sens (ce que fait très bien la langue)
- Couverture très limitée
- **Réalité cognitive, réalité conceptuelle ? ...**

Modéliser stochastiquement le sens

Des mots apparaissant dans contextes similaires ont des sens proches
(Harris, 1954)

Des mots apparaissant dans contextes similaires ont des sens proches
(Harris, 1954)

On peut étudier leur synonymie ou leur similarité

Des mots apparaissant dans contextes similaires ont des sens proches
(Harris, 1954)

On peut étudier leur synonymie ou leur similarité

Fréquences des cooccurrences → Modèles distributionnels

Des mots apparaissant dans contextes similaires ont des sens proches
(Harris, 1954)

On peut étudier leur synonymie ou leur similarité

Fréquences des cooccurrences → Modèles distributionnels

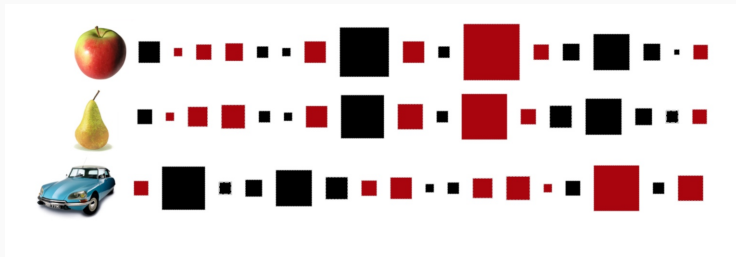
Les mots sont représentés par des valeurs réelles dans un espace avec peu de dimensions

Des mots apparaissant dans contextes similaires ont des sens proches
(Harris, 1954)

On peut étudier leur synonymie ou leur **similarité**

Fréquences des cooccurrences → Modèles distributionnels

Les mots sont représentés par des valeurs réelles dans un espace avec peu de dimensions



L'intérêt est d'avoir des mots/sens proches dans l'espace de représentation

L'intérêt est d'avoir des mots/sens proches dans l'espace de représentation

- Score de similarité entre deux mots (cosinus)
- Voisins de chat : chien, chaton, lapin, oiseau etc.

L'intérêt est d'avoir des mots/sens proches dans l'espace de représentation

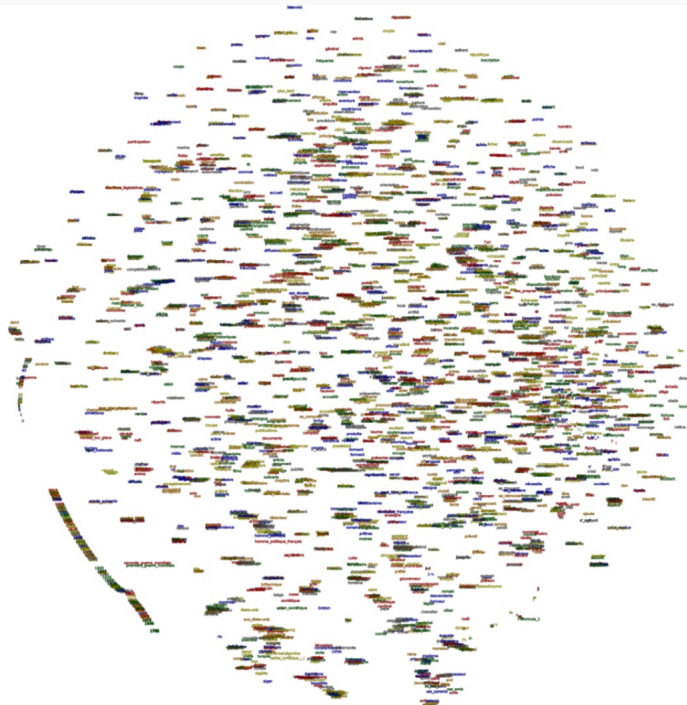
- Score de similarité entre deux mots (cosinus)
- Voisins de chat : chien, chaton, lapin, oiseau etc.
- Basé sur le comptage : extraire les fréquences de cooccurrence d'un corpus
- Basé sur la prédiction : induire des paramètres basés sur la prédiction du mot correct

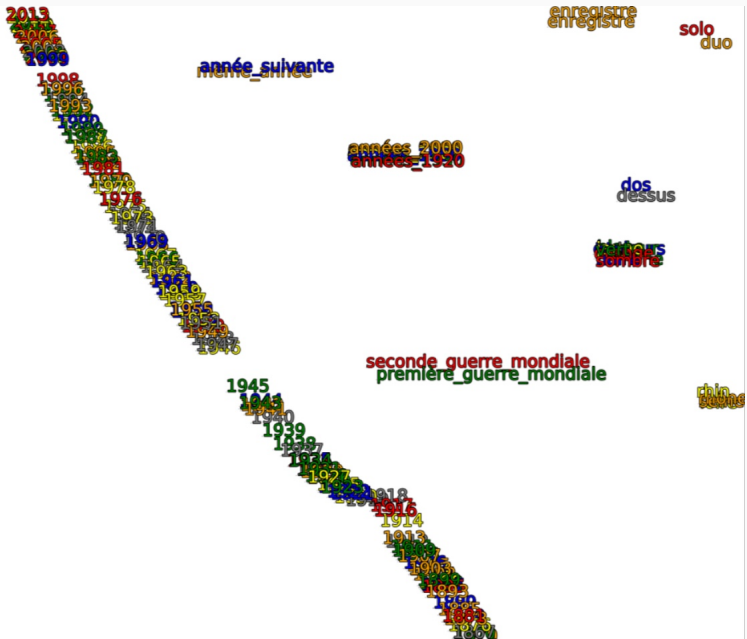
L'intérêt est d'avoir des mots/sens proches dans l'espace de représentation

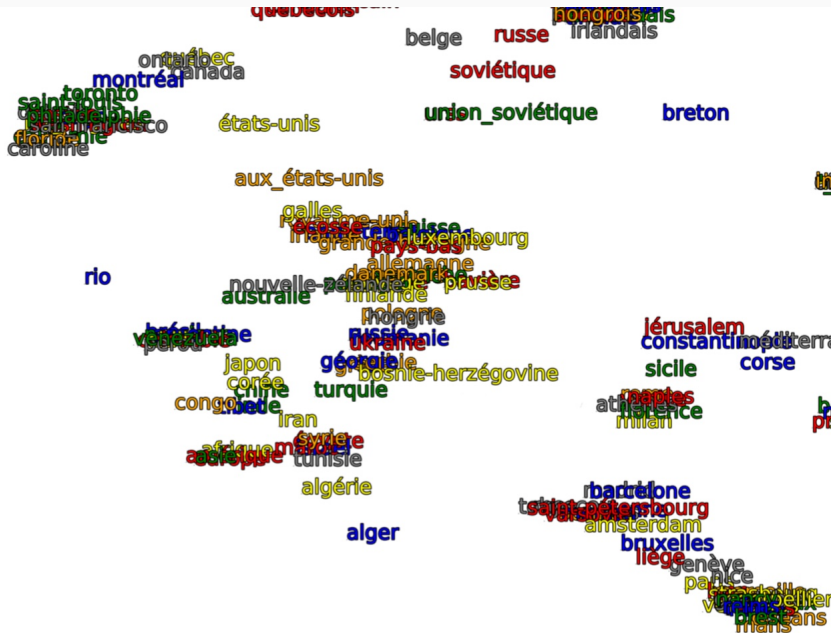
- Score de similarité entre deux mots (cosinus)
- Voisins de chat : chien, chaton, lapin, oiseau etc.
- Basé sur le comptage : extraire les **fréquences de cooccurrence** d'un corpus
- Basé sur la prédiction : induire des paramètres basés sur la **prédiction du mot correct**

Au départ, deux méthodes différentes, où l'approche fondée sur la prédiction donnait de meilleurs résultats

Par la suite, preuves qu'elles sont réellement équivalentes ou assez semblables







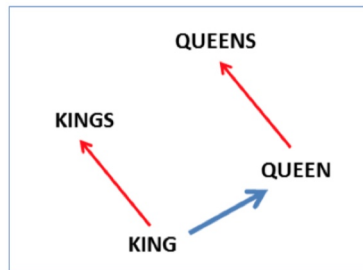
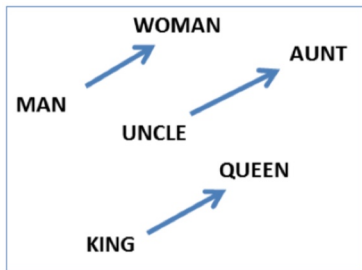


Calcul de l'analogie

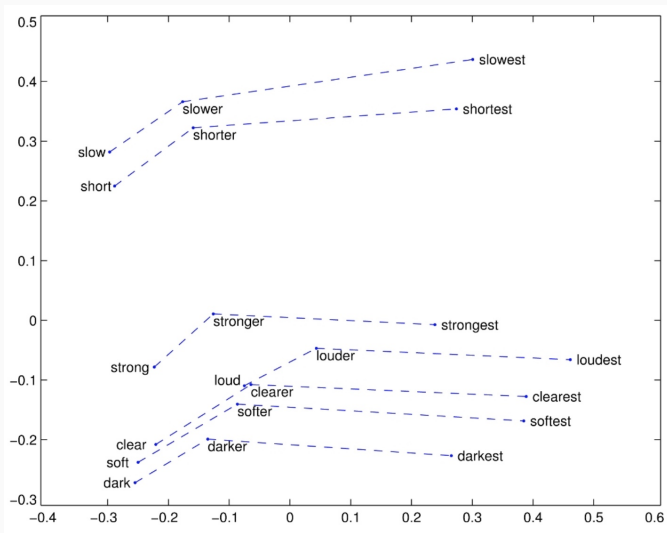
a	is to	b	as	c	is to	d
homme		roi	-	femme		reine
autriche		vienne	-	allemagne		berlin
écrivain		livre	-	poète		poème
france		Emmanuel Macron	-	états-unis		Trump
droite		LR	-	gauche		ps

Calcul de l'analogie

a	is to	b	as	c	is to	d
homme		roi	-	femme		reine
autriche		vienne	-	allemagne		berlin
écrivain		livre	-	poète		poème
france		Emmanuel Macron	-	états-unis		Trump
droite		LR	-	gauche		ps



Plongements lexicaux



- Objectif de l'entraînement
- Choix des contextes (sac de mots, syntaxe, corpus parallèle)
- Taille de la fenêtre (petite, phrase, paragraphe, dynamique)
- Dirigé
- Lemmatisation, normalisation, mots d'arrêt. . .

- Les plongements de mots sont souvent utilisés pour l'entraînement
- Non supervisé : basé sur du texte brut (donc possibilité d'utiliser de nombreuses données)
- Algorithmes rapides et disponibles
- Le modèle peut être raffiné pour une tâche

En pratique

- Les plongements de mots sont souvent utilisés pour l'entraînement
- Non supervisé : basé sur du texte brut (donc possibilité d'utiliser de nombreuses données)
- Algorithmes rapides et disponibles
- Le modèle peut être raffiné pour une tâche



- day + night =



- flying + sailing =



- bowl + box =



- box + bowl =



Pour

- Faciles à manipuler (si on les entraîne sur des corpus adéquats)
- **Bons résultats rapidement** (bonne couverture)
- Adapté pour le Deep Learning

Contre

- **Difficile d'interpréter les résultats**
- Difficultés de contrôler les paramètres du modèle

Compositionnalité dans les réseaux de neurones

Travaux récents de FAIR sur la compositionnalité dans les réseaux de neurones (!!)

Compositionnalité systématique

Brenden Lake and Marco Baroni. Generalization without systematicity : On the compositional skills of sequence-to-sequence recurrent networks. ICML 2018

- Walk
- Walk twice
- Run
- Run twice

Compositionnalité systématique

Brenden Lake and Marco Baroni. Generalization without systematicity : On the compositional skills of sequence-to-sequence recurrent networks. ICML 2018

- Walk
- Walk twice
- Run
- Run twice
- Dax

Compositionnalité systématique

Brenden Lake and Marco Baroni. Generalization without systematicity : On the compositional skills of sequence-to-sequence recurrent networks. ICML 2018

- Walk
- Walk twice
- Run
- Run twice
- Dax
- Dax twice

Compositionnalité systématique

Brenden Lake and Marco Baroni. Generalization without systematicity : On the compositional skills of sequence-to-sequence recurrent networks. ICML 2018

- Walk
- Walk twice
- Run
- Run twice
- Dax
- Dax twice

$$[[X \text{ twice}]] = [[X]][[X]]$$

$$[[dax]] = \text{realiser l'action dax}$$

TRAINING PHASE

walk
WALK

jump after walk
WALK JUMP

walk and jump left
WALK LTURN JUMP

run thrice
RUN RUN RUN

run around right
RTURN RUN RTURN RUN
RTURN RUN RTURN RUN

look right and
walk left
RTURN LOOK
LTURN WALK

walk and run
RUN WALK

TEST TIME

jump around
and run

Décomposition des commandes

Commandes de base :

- run → RUN
- walk → WALK
- turn left → LTURN

Modificateurs :

- walk left → LTURN WALK
- run twice → RUN RUN

Conjonctions :

- walk left and run twice → LTURN WALK RUN RUN
- run twice after walk left → RUN RUN LTURN WALK

Simplifications :

- Pas d'ambiguïté de porté ("walk and [run twice]")
- Pas de recursion ("walk and run" vs **"walk and run and walk")

- Entraînement d'un modèle sequence-to-sequence RNN sur 100k commandes et les séquences correspondantes
- Pour le test, uniquement des nouvelles commandes
- Chaque commande de test n'est utilisée qu'une fois
- Le modèle doit être performant immédiatement
- 2-layer LSTM avec 200 unités par niveau, pas d'attention, 0.5 dropout

Expérience 1 : Répartition aléatoire des données d'entraînement et de test

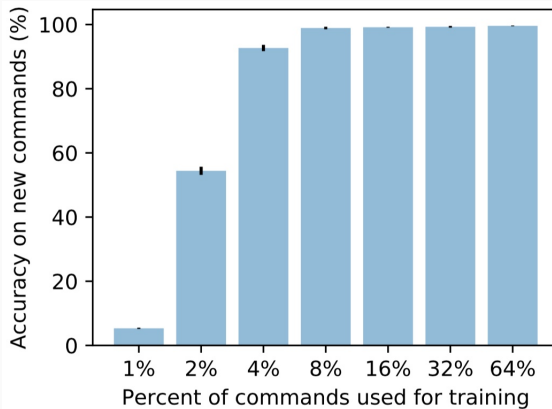
Dans les données d'entraînement :

- look around left twice
- look around left twice and turn left
- jump right twice
- run twice and jump right twice

Dans le test :

- look around left twice and jump right twice

Expérience 1 : ça marche



Expérience 2 : découpage par le nombre d'action

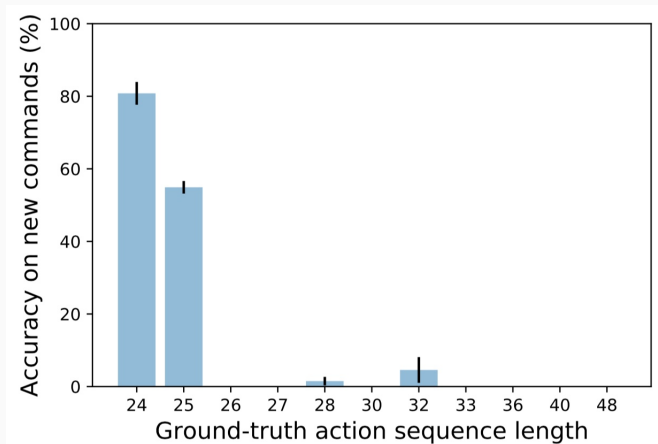
Entraînement sur les commandes avec les séquences d'actions les plus courtes (max 22 actions)

- jump around left twice (16 actions)
- walk opposite right thrice (9 actions)
- jump around left twice and walk opposite right twice (22 actions)

Test sur les commandes avec les séquences d'actions les plus longues (de 24 à 48 actions)

- jump around left twice and walk opposite right thrice (25 actions)

Expérience 2 : ça ne marche pas (du tout)



Expérience 3 : généralisation de la composition de commande de base (“dax”)

Les données d'entraînement contiennent toutes les commandes possibles avec “run”, “walk”, “look”, “turn left”, “turn right”

- “run”, “run twice”, “turn left and run opposite thrice”, “walk after run”, ...

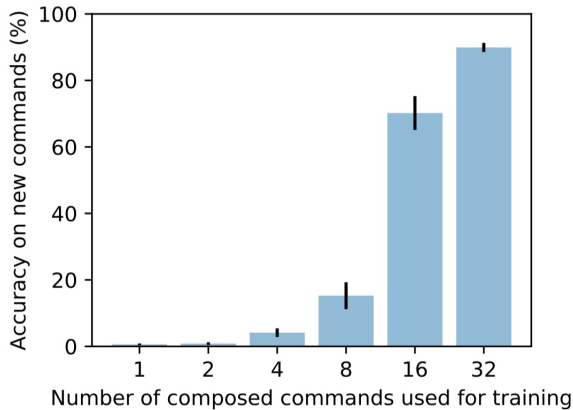
Mais seulement un petit ensemble de commandes composées avec “jump” :

- “jump”, “jump left”, “run and jump”, “jump around twice”

Le système est testé sur toutes les autres commandes “jump”

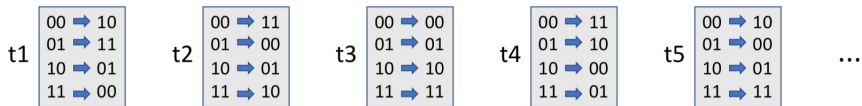
- jump twice
- jump left and run opposite thrice
- walk after jump

Expérience 3 : beaucoup de données sont nécessaires !



Problème de consultation de table

Memorize or generalize ? Searching for a compositional RNN in a haystack, Adam Liska,
Germán Kruszewski and Marco Baroni, AEGAP Workshop 2018



Problème de consultation de table

Memorize or generalize ? Searching for a compositional RNN in a haystack, Adam Liska,
Germán Kruszewski and Marco Baroni, AEGAP Workshop 2018

t1	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>11</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>00</td></tr></table>	00	→	10	01	→	11	10	→	01	11	→	00	t2	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>10</td></tr></table>	00	→	11	01	→	00	10	→	01	11	→	10	t3	<table><tr><td>00</td><td>→</td><td>00</td></tr><tr><td>01</td><td>→</td><td>01</td></tr><tr><td>10</td><td>→</td><td>10</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	00	01	→	01	10	→	10	11	→	11	t4	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>10</td></tr><tr><td>10</td><td>→</td><td>00</td></tr><tr><td>11</td><td>→</td><td>01</td></tr></table>	00	→	11	01	→	10	10	→	00	11	→	01	t5	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	10	01	→	00	10	→	01	11	→	11	...
00	→	10																																																																				
01	→	11																																																																				
10	→	01																																																																				
11	→	00																																																																				
00	→	11																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	10																																																																				
00	→	00																																																																				
01	→	01																																																																				
10	→	10																																																																				
11	→	11																																																																				
00	→	11																																																																				
01	→	10																																																																				
10	→	00																																																																				
11	→	01																																																																				
00	→	10																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	11																																																																				

$t1(00)=10$

$t3(00)=00$

$t4(t5(01))=11$

$t5(t4(01))=01$

$t2(t2(10))=00$

$t1(t4(t5(11)))=11$

$t1(t5(t1(10)))=10$

Problème de consultation de table

Memorize or generalize ? Searching for a compositional RNN in a haystack, Adam Liska,
Germán Kruszewski and Marco Baroni, AEGAP Workshop 2018

t1	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>11</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>00</td></tr></table>	00	→	10	01	→	11	10	→	01	11	→	00	t2	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>10</td></tr></table>	00	→	11	01	→	00	10	→	01	11	→	10	t3	<table><tr><td>00</td><td>→</td><td>00</td></tr><tr><td>01</td><td>→</td><td>01</td></tr><tr><td>10</td><td>→</td><td>10</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	00	01	→	01	10	→	10	11	→	11	t4	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>10</td></tr><tr><td>10</td><td>→</td><td>00</td></tr><tr><td>11</td><td>→</td><td>01</td></tr></table>	00	→	11	01	→	10	10	→	00	11	→	01	t5	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	10	01	→	00	10	→	01	11	→	11	...
00	→	10																																																																				
01	→	11																																																																				
10	→	01																																																																				
11	→	00																																																																				
00	→	11																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	10																																																																				
00	→	00																																																																				
01	→	01																																																																				
10	→	10																																																																				
11	→	11																																																																				
00	→	11																																																																				
01	→	10																																																																				
10	→	00																																																																				
11	→	01																																																																				
00	→	10																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	11																																																																				

$t1(00)=10$

$t3(00)=00$

$t4(t5(01))=11$

$t5(t4(01))=01$

$t2(t2(10))=00$

$t1(t4(t5(11)))=11$

$t1(t5(t1(10)))=10$

Aucune intelligence, les données sont mémorisées, et les possibilités de composition sont infinies

Test de la généralisation de la composition

t1

00	→	10
01	→	11
10	→	01
11	→	00

t2

00	→	11
01	→	00
10	→	01
11	→	10

t3

00	→	00
01	→	01
10	→	10
11	→	11

t4

00	→	11
01	→	10
10	→	00
11	→	01

t5

00	→	10
01	→	00
10	→	01
11	→	11

...

Entraînement #1 : consultation simple des tables

t1 :00.10. t4 :10.00. t301.01. ...

Test de la généralisation de la composition

t1	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>11</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>00</td></tr></table>	00	→	10	01	→	11	10	→	01	11	→	00	t2	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>10</td></tr></table>	00	→	11	01	→	00	10	→	01	11	→	10	t3	<table><tr><td>00</td><td>→</td><td>00</td></tr><tr><td>01</td><td>→</td><td>01</td></tr><tr><td>10</td><td>→</td><td>10</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	00	01	→	01	10	→	10	11	→	11	t4	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>10</td></tr><tr><td>10</td><td>→</td><td>00</td></tr><tr><td>11</td><td>→</td><td>01</td></tr></table>	00	→	11	01	→	10	10	→	00	11	→	01	t5	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	10	01	→	00	10	→	01	11	→	11	...
00	→	10																																																																				
01	→	11																																																																				
10	→	01																																																																				
11	→	00																																																																				
00	→	11																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	10																																																																				
00	→	00																																																																				
01	→	01																																																																				
10	→	10																																																																				
11	→	11																																																																				
00	→	11																																																																				
01	→	10																																																																				
10	→	00																																																																				
11	→	01																																																																				
00	→	10																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	11																																																																				

Entraînement #1 : consultation simple des tables

t1 :00.10. t4 :10.00. t301.01. ...

Entraînement #2 : consultation simple et composée des tables

ct1t4 :00 :00. t3 :10.10. ct5t5 :01.10. ...

Test de la généralisation de la composition

t1	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>11</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>00</td></tr></table>	00	→	10	01	→	11	10	→	01	11	→	00	t2	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>10</td></tr></table>	00	→	11	01	→	00	10	→	01	11	→	10	t3	<table><tr><td>00</td><td>→</td><td>00</td></tr><tr><td>01</td><td>→</td><td>01</td></tr><tr><td>10</td><td>→</td><td>10</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	00	01	→	01	10	→	10	11	→	11	t4	<table><tr><td>00</td><td>→</td><td>11</td></tr><tr><td>01</td><td>→</td><td>10</td></tr><tr><td>10</td><td>→</td><td>00</td></tr><tr><td>11</td><td>→</td><td>01</td></tr></table>	00	→	11	01	→	10	10	→	00	11	→	01	t5	<table><tr><td>00</td><td>→</td><td>10</td></tr><tr><td>01</td><td>→</td><td>00</td></tr><tr><td>10</td><td>→</td><td>01</td></tr><tr><td>11</td><td>→</td><td>11</td></tr></table>	00	→	10	01	→	00	10	→	01	11	→	11	...
00	→	10																																																																				
01	→	11																																																																				
10	→	01																																																																				
11	→	00																																																																				
00	→	11																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	10																																																																				
00	→	00																																																																				
01	→	01																																																																				
10	→	10																																																																				
11	→	11																																																																				
00	→	11																																																																				
01	→	10																																																																				
10	→	00																																																																				
11	→	01																																																																				
00	→	10																																																																				
01	→	00																																																																				
10	→	01																																																																				
11	→	11																																																																				

Entraînement #1 : consultation simple des tables

t1 :00.10. t4 :10.00. t301.01. ...

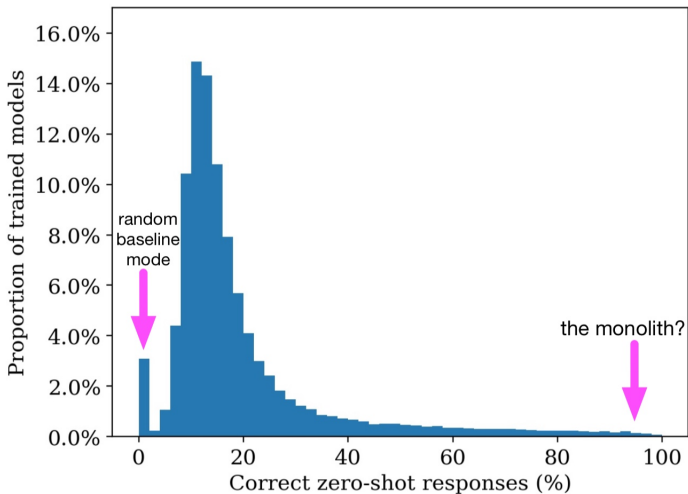
Entraînement #2 : consultation simple et composée des tables

ct1t4 :00 :00. t3 :10.10. ct5t5 :01.10. ...

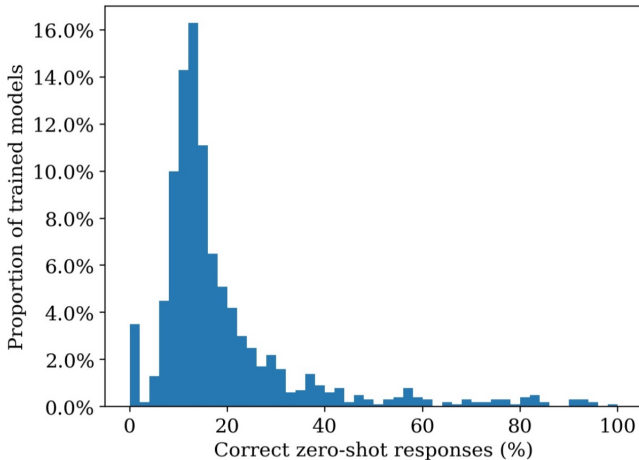
Test : consultation composées avec de nouvelles données

ct1t4 :01 :01. ct5t5 :00.01. ct3t2 :10.01.

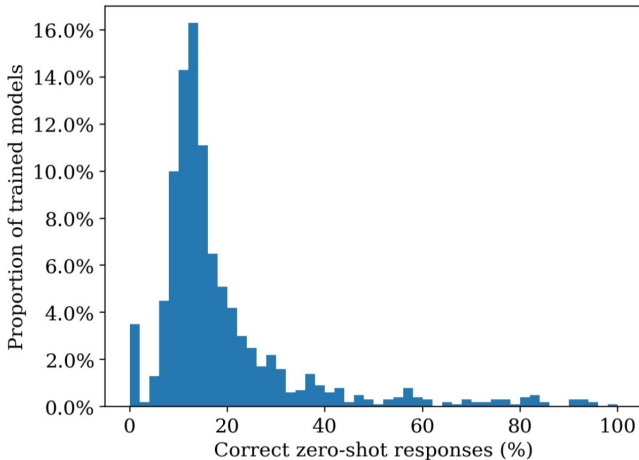
- Réseaux de neurones avec deux niveaux cachés, architecture théoriquement apte pour la composition (Recurrent 60-unit LSTM layer, 10-unit sigmoid layer)
- 1M d'exemples pour les entraînement des phases #1 et #2
- 128 données conservées pour le test (2 par composition de table de premier ordre possible)
- Entraînement standard : rétro-propagation des paramètres, mise à jour avec descente stochastique du gradient (mises à jour parallèles à partir de 40 processeurs)
- Expérience répétée 50k fois avec des initialisations aléatoires



Même initialisation, autre apprentissage



Et en cachant les parties compositionnelles (!!)



Conclusion

- L'approche **statistique** de la sémantique de **la langue** donne de (très) **bons résultats**
- En particulier grâce aux **réseaux de neurones** qui sont agnostiques sur la tâche

- L'approche **statistique** de la sémantique de **la langue** donne de (très) **bons résultats**
- En particulier grâce aux **réseaux de neurones** qui sont agnostiques sur la tâche
- Mais leur généralisation **ne capte pas systématiquement les propriétés compositionnelles**
- La **compositionnalité** reste complexe pour les humains

- L'approche **statistique** de la sémantique de **la langue** donne de (très) **bons résultats**
- En particulier grâce aux **réseaux de neurones** qui sont agnostiques sur la tâche
- Mais leur généralisation **ne capte pas systématiquement les propriétés compositionnelles**
- La **compositionnalité** reste complexe pour les humains
- Un enjeu important est celui de comprendre/modéliser la composition dans ces approches

- L'approche **statistique** de la sémantique de **la langue** donne de (très) **bons résultats**
- En particulier grâce aux **réseaux de neurones** qui sont agnostiques sur la tâche
- Mais leur généralisation **ne capte pas systématiquement les propriétés compositionnelles**
- La **compositionnalité** reste complexe pour les humains
- Un enjeu important est celui de comprendre/modéliser la composition dans ces approches
- ou de simuler la compositionnalité par composition de différents réseaux

- L'approche **statistique** de la sémantique de **la langue** donne de (très) **bons résultats**
- En particulier grâce aux **réseaux de neurones** qui sont agnostiques sur la tâche
- Mais leur généralisation **ne capte pas systématiquement les propriétés compositionnelles**
- La **compositionnalité** reste complexe pour les humains
- Un enjeu important est celui de comprendre/modéliser la composition dans ces approches
- ou de simuler la compositionnalité par composition de différents réseaux
- ou d'utiliser en entrée des données rendant compte de ces phénomènes

- On calcule relativement bien sur la langue
- mais on comprend plutôt superficiellement
- ou très précisément de choses très compliquées (et uniquement celles-là)

Merci !